Available online at www.sciencedirect.com
SCIENCE DIRECT®
Advances in Water Resources 26 (2003) 1177–1188
Advances in Water Resources
www.elsevier.com/locate/advwatres

On importance sampling Monte Carlo approach to uncertainty analysis for flow and transport in porous media

Zhiming Lu *, Dongxiao Zhang

Hydrology, Geochemistry, and Geology Group (EES-6), MS T003, Los Alamos National Laboratory, Los Alamos, NM 87545, USA

Received 17 April 2003; received in revised form 29 July 2003; accepted 30 July 2003

Abstract

In this study we introduce a new approach named importance sampling or quick simulations. The method has been extensively used in communication theory in estimating probability of rare events. The basic idea behind importance sampling techniques is that certain values of the input random variables (or vectors) have more important impact on the parameters being estimated than others, and if these "important" values are sampled more frequently than others, i.e., sampled from a biased density function, the variance of the estimator can be reduced. The outputs from simulations are then weighted to correct such biased sampling. Two illustrative examples are given to show the general procedure of the importance sampling approach as well as its applicability to subsurface flow and transport problems. In one example we estimated the mean and variance of hydraulic head for one-dimensional flow, and in the other we estimated the probability of a particle's travel time t less than a given critical value T. In both examples, we compared results from analytical solutions, the conventional Monte Carlo (CMC) simulations, and the importance sampling approach. It is shown that when an importance density function is chosen appropriately, importance sampling techniques may be many orders of magnitude more efficient than the CMC simulations and have a great potential in simulating subsurface flow and transport.
Published by Elsevier Ltd.

Keywords: Monte Carlo simulation; Importance sampling; Heterogeneity; Flow and transport; Variance reduction; Variance estimation

1. Introduction

It is well known that geological formations are ubiquitously heterogeneous, and the equations describing flow and transport in these formations are stochastic. Stochastic approaches to flow and transport in heterogeneous porous media have been extensively studied in the last two decades, and many stochastic models have been developed [4,6,19]. These models can be classified into two groups: Monte Carlo simulations and moment-equation approaches. Monte Carlo simulation is a conceptually straightforward approach to solve these stochastic partial differential equations by generating a large number of equally likely random realizations of the parameter fields, solving deterministic flow and transport equations for each realization, and averaging the results over all realizations to obtain sample moments of the solution. This approach has the advantages of applying to a very broad range of both linear and nonlinear flow and transport problems, but has a number of potential drawbacks.

A major disadvantage of the Monte Carlo method, among others, is the requirement for large computational effort. To properly resolve high frequency space–time fluctuations in random parameters, it is necessary to employ fine numerical grids in space–time. Therefore, computational effort for each realization is usually large, especially if both physical and chemical heterogeneities, as well as uncertainties in initial and boundary conditions, are considered. To ensure sample output moments converge to their theoretical ensemble values, a large number of Monte Carlo simulations are often required (typically a few thousand realizations, depending on the degree of medium heterogeneity), which dramatically increases the computation burden.

One alternative to Monte Carlo simulations is an approach based on moment equations, the essence of which is to derive a system of deterministic partial differential equations governing the first two moments (mean and covariance), and then solve them analytically or numerically [3,4,7,11,17–20].

* Corresponding author. Tel.: +1-505-665-2126.
E-mail address: zhiming@lanl.gov (Z. Lu).

0309-1708/$ - see front matter Published by Elsevier Ltd.
doi:10.1016/S0309-1708(03)00106-4

Though the moment-equation-based approach (based on perturbation methods) in many cases works well for relatively large variations in the medium properties [10,16,21], this approach in general is restricted to small variabilities of medium properties. Although the moment equation approach may render accurate solutions with coarser numerical grids (of fewer nodes) than the Monte Carlo method, it must solve covariance equations as many times as the number of nodes. Thus, this approach can also be computationally demanding, in particular, for large-size problems.

In this study we introduce a new approach named importance sampling or quick simulations for the purpose of solving problems of flow and transport in random porous media. The importance sampling method is one of variance reduction techniques and it has been extensively used in communication theory in simulating rare events [1,9,13]. The basic idea behind importance sampling techniques is that certain values of the input random variables (or vectors) have more important impact on the quantities being estimated than others, and if these "important" values are sampled more frequently, i.e., sampled from a biased density function, the variance of the estimator can be reduced. The outputs from simulations are then weighted to correct the bias caused by sampling from the biased density function. A detailed mathematical framework for importance sampling is given in [8] and a thorough review on this topic can be found in [14].

The organization of this paper is as follows. In Section 2, we review the principle of importance sampling techniques in estimating mean quantities and develop for the first time a similar principle for estimating their associated variances. We also discuss in this section the choice of importance density functions. Two examples are given in Section 3 to illustrate the general procedure of the importance sampling approach and its applicability to subsurface flow and transport problems. In one example we estimated the mean and variance of hydraulic head for one-dimensional flow, and in the other we estimated the probability of a particle's travel time t less than a given critical value T. In both examples, we compared results from analytical solutions, the conventional Monte Carlo (CMC) simulations, and the importance sampling approach. It is shown that in the cases where the importance density function is chosen appropriately, the importance sampling technique may be many orders of magnitude more efficient than the CMC simulations.

2. Principle of importance sampling techniques

The purpose of the importance sampling techniques is to obtain accurate estimate of quantities (such as mean hydraulic head and head variance) with fewer samples than required in the CMC method. There are two major steps involved in the importance sampling techniques. One is distortion of the original input process. Instead of taking samples from the original probability density function (pdf), samples are taken from some other pdf, called importance density functions, such that some "important" regions of the sample space get more samples. The fundamental issue in implementing the importance sampling techniques is the choice of biased importance density functions. The other is correction of the distortion by averaging the output from different samples (realizations) using weights that are related to the distortion, such that the mean of the quantity being estimated is preserved.

Let $\mathbf{Y}$ denote a random variable (or vector) with a (joint) pdf $f(\mathbf{y})$. Here $\mathbf{Y}$ may represent properties of porous media, such as log hydraulic conductivity, at discretized grid nodes. We wish to derive the moments of function $h = g(\mathbf{Y})$, where g is a specified, deterministic function (operator). For example, g can be a numerical solver that maps a log hydraulic conductivity field $\mathbf{Y}$ into a head or velocity field h, under some given boundary and initial conditions.

2.1. Mean estimation

By definition, the mean and variance of h can be written as

$$\mu_h = E[h] = E[g(\mathbf{Y})] = \int_\Omega g(\mathbf{y}) f(\mathbf{y})\,\mathrm{d}\mathbf{y}, \tag{1}$$

$$\sigma_h^2 = E\{[g(\mathbf{Y}) - \mu_h]^2\} = \int_\Omega [g(\mathbf{y}) - \mu_h]^2 f(\mathbf{y})\,\mathrm{d}\mathbf{y} = \int_\Omega g^2(\mathbf{y}) f(\mathbf{y})\,\mathrm{d}\mathbf{y} - \mu_h^2, \tag{2}$$

where Ω is a probability space, and E represents statistical expectation. Note that both μ_h and σ_h^2 depend on the problem itself and are independent of the way one estimates them. To estimate the mean from Eq. (1) using the CMC method, one samples randomly a sequence of $\mathbf{y}_i$, $i = 1, 2, \ldots, N$, from the density function $f(\mathbf{y})$ and computes the sample mean

$$\bar{h}_N = \frac{1}{N}\sum_{i=1}^{N} g(\mathbf{y}_i). \tag{3}$$

Since $\mathbf{y}_i$, $i = 1, 2, \ldots, N$ are independent identically distributed random variables, it can be shown that $E[\bar{h}_N] = \mu_h$, i.e., $\bar{h}_N$ is an unbiased estimator of μ_h, and the variance of the estimator $\bar{h}_N$ is $\overline{\sigma^2_{\mathrm{MC}}} = \sigma_h^2/N$. Because σ_h^2 is unknown, it is usually estimated using

$$S^2_{h,N} = \frac{1}{N}\sum_{i=1}^{N} g^2(\mathbf{y}_i) - \bar{h}_N^2. \tag{4}$$

By the Central Limit Theorem, in the limit $N \to \infty$, the probability density of $(\bar{h}_N - \mu_h)/(\sigma_h/\sqrt{N})$ tends to a standard Gaussian distribution $N(0,1)$. Thus we expect $\bar{h}_N$ to lie within $\pm\sigma_h/\sqrt{N}$ around μ_h with the probability of 68%, if N is sufficiently large. From this, we can estimate the minimum number of simulations required to obtain the mean estimation with a given accuracy. Let $N_{\epsilon,\mathrm{MC}}$ denote the minimum number of simulations required to obtain a $100 \times \epsilon\%$ precision, then

$$N_{\epsilon,\mathrm{MC}} = \lceil \sigma_h^2/\mu_h^2\epsilon^2 \rceil, \tag{5}$$

where $\lceil x \rceil$ is the least integer $\geqslant x$. For instance, if we need the estimation to be within 10% around the exact value, the minimum number of required simulations should be $N_{0.1,\mathrm{MC}} = \lceil \sigma_h^2/\mu_h^2\epsilon^2 \rceil = \lceil 100\sigma_h^2/\mu_h^2 \rceil$.

Because the statistical error $\sigma_h/\sqrt{N}$ is inversely proportional to $\sqrt{N}$, if we wish to reduce it by a factor of two we have to increase the sample size by a factor of four. Certainly, the rate of convergence of the CMC method is rather slow. It is desirable if we can design some ways to reduce the estimation variance more rapidly as N increases.

The importance sampling technique is one of such ways that reduce the estimation variance (called variance reduction techniques) and thus reduce the statistical error much faster than the CMC techniques. The efficiency of the importance sampling techniques depends on the selected importance density functions. The essence of importance sampling is to avoid taking samples $\mathbf{y}$ in regions where the impact of the value of the function $g(\mathbf{y})$ to the quantity being estimated is negligible but to concentrate on (important) regions where the impact is large. This operation will inevitably introduce bias (or, distortion), which is to be corrected by weighting the sample values appropriately.

Suppose we sample $\mathbf{y}_i$, $i = 1, 2, \ldots, N$, from an importance density function $f_1(\mathbf{y})$ rather than the original density function $f(\mathbf{y})$, where $f_1(\mathbf{y})$ is zero only if $f(\mathbf{y})$ is zero. To preserve the mean (i.e., to correct the bias) one has to modify the original score function $g(\mathbf{y})$. A modified score function $H = g_1(\mathbf{Y})$ is defined as

$$g_1(\mathbf{y}) = g(\mathbf{y})w_1(\mathbf{y}), \tag{6}$$

where $w_1(\mathbf{y}) = f(\mathbf{y})/f_1(\mathbf{y})$ is called a weight function. The expectation of H under density $f_1(\mathbf{y})$ can be determined

$$\begin{aligned}\mu_H = E[H] = E[g_1(\mathbf{Y})] &= \int_\Omega g_1(\mathbf{y})f_1(\mathbf{y})\,\mathrm{d}\mathbf{y} \\ &= \int_\Omega g(\mathbf{y})\frac{f(\mathbf{y})}{f_1(\mathbf{y})}f_1(\mathbf{y})\,\mathrm{d}\mathbf{y} = \int_\Omega g(\mathbf{y})f(\mathbf{y})\,\mathrm{d}\mathbf{y} = \mu_h,\end{aligned} \tag{7}$$

which means that the mean remains the same though the samples are taken biasedly from the importance density function $f_1(\mathbf{y})$. The variance H under density $f_1(\mathbf{y})$ is

$$\begin{aligned}\sigma_H^2 = E\{[g_1(\mathbf{Y}) - \mu_H]^2\} &= \int_\Omega [g_1(\mathbf{y}) - \mu_H]^2 f_1(\mathbf{y})\,\mathrm{d}\mathbf{y} \\ &= \int_\Omega g_1^2(\mathbf{y})f_1(\mathbf{y})\,\mathrm{d}\mathbf{y} - \mu_H^2 \\ &= \int_\Omega \left[\frac{f(\mathbf{y})}{f_1(\mathbf{y})}\right] g^2(\mathbf{y})f(\mathbf{y})\,\mathrm{d}\mathbf{y} - \mu_H^2.\end{aligned} \tag{8}$$

From (7) one can construct an estimator for μ_h based on samples $\mathbf{y}_i$, $i = 1, 2, \ldots, N$:

$$\overline{H}_N = \frac{1}{N}\sum_{i=1}^{N} g_1(\mathbf{y}_i) = \frac{1}{N}\sum_{i=1}^{N} g(\mathbf{y}_i)w_1(\mathbf{y}_i), \tag{9}$$

i.e., the contribution of sample $\mathbf{y}_i$ in $\overline{H}_N$ is weighted by $w_1(\mathbf{y}_i)$ and the bias due to sampling from the biased importance density function has been corrected. The variance of the estimator $\overline{H}_N$ is $\overline{\sigma_{\mathrm{IS}}^2} = \sigma_H^2/N$, where σ_H^2 can be estimated using

$$S_{H,N}^2 = \frac{1}{N}\sum_{i=1}^{N} g^2(\mathbf{y}_i)w_1^2(\mathbf{y}_i) - \overline{H}_N^2. \tag{10}$$

Similar to equation (5), the minimum number of simulations required to obtain an estimation of $E[g]$ with a $100 \times \epsilon\%$ precision is

$$N_{\epsilon,\mathrm{IS}} = \lceil \sigma_{\mathrm{IS}}^2/\mu_H^2\epsilon^2 \rceil. \tag{11}$$

The ratio of variances $\gamma = \overline{\sigma_{\mathrm{MC}}^2}/\overline{\sigma_{\mathrm{IS}}^2} = N_{\epsilon,\mathrm{MC}}/N_{\epsilon,\mathrm{IS}}$ is a measure of the efficiency (performance) of the importance sampling, which depends on the choice of the importance density function. Here $\overline{\sigma_{\mathrm{MC}}^2} = \sigma_h^2/N$ is the estimation variance of the CMC method. It should be noted by comparison between (8) and (2) that if we choose the importance density function $f_1(\mathbf{y})$ properly such that the weight function $w_1(\mathbf{y})$ on average is substantially less than unity, σ_H^2 and thus the variance of the estimator $\overline{\sigma_{\mathrm{IS}}^2} = \sigma_H^2/N$ will be reduced. This means that, comparing to the CMC method, sampling from the importance density function may allow us to estimate the mean with a small sample size for a given accuracy, or with a much better accuracy for a given sample size. In fact, as will be shown later, the Monte Carlo approach based on importance sampling techniques makes it possible to solve some problems that cannot be solved by the CMC simulation.

We should emphasize here that uncertainties associated with modeling applications can be classified as "reducible" and "irreducible". Natural uncertainty is "inherited" or irreducible, while data and model uncertainties contain both reducible and irreducible components. By the term "variance reduction" we refer to reduction of estimation variance, i.e., the variance of the estimator or model uncertainty, not the variance inherited from uncertainties in system parameters. For instance, in evaluating an integral $\int_a^b f(x)\,\mathrm{d}x$ numerically, one may discretize interval $[a, b]$ using N points that can

be either uniformly or randomly distributed in $[a,b]$ and then calculate $\sum f(x_i)\Delta x_i$. The accuracy of this estimation depends on the number of points, N, and the distribution of these points in $[a,b]$. A good estimator may require less computation time, converge faster to the true value, and thus have a smaller estimation variance. Note that here the variance caused by uncertainties is not involved, because the value of the integral is a deterministic quantity. In general, the variance due to parameter uncertainties cannot be reduced unless more information about system parameters is added. For example, if we are interested in estimating mean head for given uncertainty of hydraulic conductivity, the variance introduced by the spatial variability of hydraulic conductivity cannot be reduced unless more information, such as more measurements on hydraulic conductivity, is provided. Of course, we may be able to design a good estimator (such as importance sampling) such that the estimator converges to the true value faster, i.e., reducing the variance of the estimator.

2.2. Variance estimation

Unlike in communication systems where the mean, bit error rate, is the main quantity to be estimated, in hydrology it is often needed to estimate the variance associated with the mean prediction. Because importance sampling techniques do not preserve variance (i.e., σ_H^2 is in general not equal to σ_h^2), if we want to estimate variance σ_h^2, we will have to re-formulate the expression for variance. From Eq. (2) we have

$$\sigma_h^2 = E\{[g(\mathbf{Y}) - \mu_h]^2\} = \int_\Omega g^2(\mathbf{y}) f(\mathbf{y})\,d\mathbf{y} - \mu_h^2 = \int_\Omega \frac{f_1(\mathbf{y})}{f(\mathbf{y})} g_1^2(\mathbf{y}) f_1(\mathbf{y})\,d\mathbf{y} - \mu_H^2, \tag{12}$$

where μ_h has been replaced by μ_H because they are the same. In Eq. (12), for any $\mathbf{y}$ such that $f(\mathbf{y}) \equiv 0$, from Eq. (6), $g_1(\mathbf{y})$ must be zero. From Eq. (12), if we take samples $\mathbf{y}_i$, $i = 1,2,\ldots,N$, from an importance density function $f_1(\mathbf{y})$, then the original variance σ_h^2 may be estimated using the following expression:

$$S_{h,N}^2 = \frac{1}{N}\sum_{i=1}^{N} \frac{f_1(\mathbf{y}_i)}{f(\mathbf{y}_i)} g_1^2(\mathbf{y}_i) - \overline{H}_N^2 = \frac{1}{N}\sum_{i=1}^{N} g^2(\mathbf{y}_i) w_1(\mathbf{y}_i) - \overline{H}_N^2. \tag{13}$$

A comparison with Eq. (10) reveals that the reduction of variance $S_{H,N}^2$ in Eq. (10) is due to the fact that $g^2(\mathbf{y})$ is weighted by $w_1^2(\mathbf{y})$ in $S_{H,N}^2$ rather than by $w_1(\mathbf{y})$ as in $S_{h,N}^2$ in Eq. (13).

It should be noted that, though Eq. (13) gives corrected estimation of variance, the rate of convergence of estimator $S_{h,N}^2$ in Eq. (13) to σ_h^2 as $n \to \infty$ may not necessarily be optimized, because the importance density function $f_1(\mathbf{y})$ discussed above is selected for the purpose of preserving the mean quantity $\mu_H = \mu_h$ while reducing the computational effort by minimizing Eq. (8).

In some special cases, as shown in case 2 in the next section, the score function $g(\mathbf{y})$ is an indicator function thus $g^2(\mathbf{y}) \equiv g(\mathbf{y})$, and from Eq. (2) we have $\sigma_h^2 = \mu_H(1-\mu_H)$. Therefore, in this case once we have estimated μ_H, the variance can be easily calculated without resorting to Eq. (13).

However, in general, from Eq. (2), if one wishes to estimate σ_h^2 using smaller number of simulations, one may need to construct a new importance density function $f_2(\mathbf{y})$ to estimate the mean of $g^2(\mathbf{y})$, just as we construct $f_1(\mathbf{y})$ to estimate the mean of $g(\mathbf{y})$. Suppose we sample $\mathbf{y}_i$, $i = 1,2,\ldots,N$, from an importance density function $f_2(\mathbf{y})$, a function to be determined. Again, $f_2(\mathbf{y})$ is zero at $\mathbf{y}$ only if $f(\mathbf{y})$ is zero. By defining $g_2(\mathbf{y}) = g(\mathbf{y})w_2(\mathbf{y})$, where $w_2(\mathbf{y}) = \sqrt{f(\mathbf{y})/f_2(\mathbf{y})}$, the last integral in Eq. (2) becomes

$$E[h^2] = E[g^2(\mathbf{Y})] = \int_\Omega g^2(\mathbf{y}) f(\mathbf{y})\,d\mathbf{y} = \int_\Omega g_2^2(\mathbf{y}) f_2(\mathbf{y})\,d\mathbf{y}, \tag{14}$$

which means that the mean of $g^2(\mathbf{y})$ under distribution $f(\mathbf{y})$ is preserved by the mean of $g_2^2(\mathbf{y})$ under distribution $f_2(\mathbf{y})$. Therefore, the variance σ_h^2 is preserved under distribution $f_2(\mathbf{y})$. From the last integral of Eq. (14), $E[h^2]$ can be estimated by

$$\overline{h^2}_N = \frac{1}{N}\sum_{i=1}^{N} g_2^2(\mathbf{y}_i) = \frac{1}{N}\sum_{i=1}^{N} g^2(\mathbf{y}_i) w_2^2(\mathbf{y}_i). \tag{15}$$

Therefore, instead of using Eq. (13) one may use the following expression to estimate σ_h^2 in Eq. (2)

$$S_{h,N}^2 = \frac{1}{N}\sum_{i=1}^{N} g^2(\mathbf{y}_i) w_2^2(\mathbf{y}_i) - \overline{H}_N^2, \tag{16}$$

where $\overline{H}_N$ is estimated from Eq. (9) using weighting function $w_1(\mathbf{y})$, and samples $\mathbf{y}_i$, $i = 1,2,\ldots,N$, are taken from the importance density $f_2(\mathbf{y})$. The latter can be selected by minimizing the variance

$$\sigma_{h^2,\mathrm{IS}}^2 = \int_\Omega [g_2^2(\mathbf{y}) - E[h^2]]^2 f_2(\mathbf{y})\,d\mathbf{y} = \int_\Omega g^4(\mathbf{y}) w_2^4(\mathbf{y}) f_2(\mathbf{y})\,d\mathbf{y} - (E[h^2])^2 = \int_\Omega \frac{f(\mathbf{y})}{f_2(\mathbf{y})} g^4(\mathbf{y}) f(\mathbf{y})\,d\mathbf{y} - (E[h^2])^2. \tag{17}$$

The variance $\sigma_{h^2,\mathrm{IS}}^2$ of estimator $S_{h,N}^2$ can be estimated from importance samples $\mathbf{y}_i$:

$$S_{h^2,\mathrm{IS}}^2 = \frac{1}{N}\sum_{i=1}^{N} g^4(\mathbf{y}_i) w_2^4(\mathbf{y}_i) - (\overline{h^2}_N)^2. \tag{18}$$

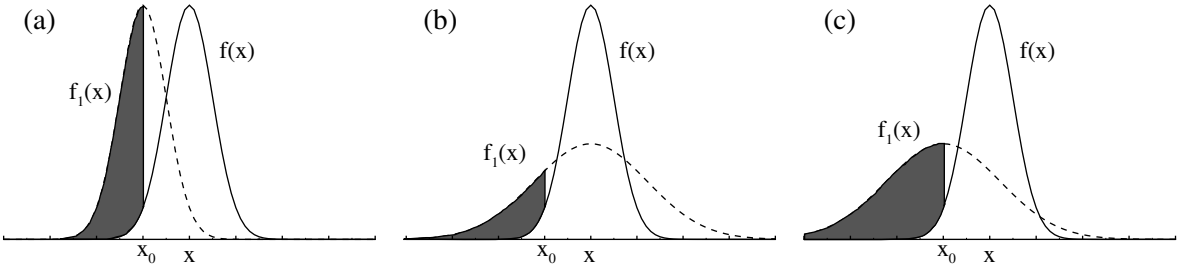

Fig. 1. Schematic diagrams showing: (a) MT, (b) VS, and (c) their combination in one dimension. Shaded area indicates the effective IS probability mass.

To our best knowledge, the importance sampling procedure for variance estimation in (16) and (17) is developed for the first time. In a special case where $g(\mathbf{y})$ is an indicator function, $g^2(\mathbf{y}) = g(\mathbf{y})$, thus $E[h^2] = E[g^2(\mathbf{y})] = E[g(\mathbf{y})] = \mu_h$. A comparison of the last integrals of Eqs. (8) and (17) reveals that in this case the importance density that minimizes Eq. (8) will also minimize Eq. (17), i.e., $f_2(\mathbf{y}) = f_1(\mathbf{y})$. In general, $f_2(\mathbf{y})$ is, however, different from $f_1(\mathbf{y})$.

In summary, to estimate a mean quantity using the importance sampling techniques, one chooses an importance density function f_1 such that σ_H^2 in Eq. (8) or $S_{H,N}^2$ in Eq. (10) is minimized, samples y from $f_1(y)$, and computes $\overline{H}_N$ in Eq. (9) as an estimate of μ_h. Similarly, to efficiently estimate variance σ_h^2, one finds f_2 such that $\sigma_{h^2,\mathrm{IS}}^2$ in (17) or $S_{h^2,\mathrm{IS}}^2$ in (18) is minimized, takes samples from f_2, and estimates σ_h^2 using $S_{h,N}^2$ in (16).

2.3. Selection of importance density functions

One important question that remains is how to choose the importance density function such that the IS estimator has a smaller variance than the CMC estimator. Here we limit our discussion on selection of $f_1(\mathbf{y})$ which minimizes Eq. (8) and reduces computational effort in estimating mean quantity μ_h. A similar procedure can also be applied to selection of $f_2(\mathbf{y})$ that minimizes Eq. (17) and reduces the number of simulations required in estimating variance σ_h^2. Ideally, if one chooses $f_1(\mathbf{y}) = g(\mathbf{y})f(\mathbf{y})/\mu_h$, then from Eq. (8) it follows that $\sigma_H^2 \equiv 0$. Though $f_1(\mathbf{y})$ defined here contains μ_h that is the quantity being estimated and thus cannot be used in practice, it does give us some clues about how to construct an importance density function.

It is seen from Eq. (8) that it is difficult to minimize σ_H^2 directly as a function of an unknown function $f_1(\mathbf{y})$. A practical alternative is to choose $f_1(\mathbf{y})$ from a family of candidates for which sampling $\mathbf{y}_i$ and evaluating $w_1(\mathbf{y})$ are relatively easy. In addition, choosing $f_1(\mathbf{y})$ from a family of candidates makes it possible in some cases to derive an analytical expression for σ_H^2.

If the original distribution is Gaussian, then the family of Gaussian distributions should be a natural choice. Two major approaches have been used in literature, i.e, variance scaling (VS) [5] and mean translation (MT) [9], or their combination. The concepts of MT and VS are illustrated in Fig. 1, where $f(x)$ is the original density function. Suppose the problem being concerned is related to a small probability $\Pr(x < x_0)$, the CMC simulation that takes samples from $f(x)$ will not be very efficient, because it is very hard to take a sample from $f(x)$ such that $x < x_0$ when x_0 is very small. By MT, one simply shifts the mean of $f(x)$ to x_0, thus the frequency to take samples from $x < x_0$ from the new density function $f_1(x)$ has been increased (Fig. 1(a)). Another way to enlarge the sampling frequency for $x < x_0$ is to increase variance (Fig. 1(b)). Fig. 1(c) illustrates the combination of MT and VS. Discussion on advantages and disadvantages of MT and VS can be found in literature [1,2,14]. In some cases, as showed by Chen et al. [2], VS in combination with optimized MT can add a degree of robustness with respect to severe nonlinearity.

It is difficult in general to analytically derive expressions for optimal parameters at which σ_H^2 in Eq. (8) is minimized, even when the form of $f(x)$ is relatively simple. In practice, the optimal parameter values can be obtained numerically. For example, suppose the original distribution is $N(\mu, \sigma^2)$, to find the optimal MT, one may choose a set of μ_i, $i = 1, 2, \ldots, m$, that spans the range of μ, take samples from $N(\mu_i, \sigma^2)$, and for each μ_i compute $S_{H,N}^2$ defined in Eq. (10). The μ_k corresponding to the minimum value of $S_{H,N}^2$ among all m values is considered as the optimal mean μ_{opt} for the optimal density function.

An alternative is an iterative approach called Spanier's technique proposed by Spanier [15] for optimizing a parameter through processing of the Monte Carlo results from a relatively small samples. Expressing the last integral in Eq. (8) as

$$M_2(\mu_i) = \int_\Omega \frac{g^2(\mathbf{y})f^2(\mathbf{y})}{f_1(\mathbf{y},\mu_i)f_1(\mathbf{y},\tilde{\mu})} f_1(\mathbf{y},\tilde{\mu})\,\mathrm{d}\mathbf{y}, \quad i = 1, 2, \ldots, m \tag{19}$$

where $\tilde{\mu}$ is a guess value of the parameter μ, and μ_i are parameter values that span the most likely range of μ. Now start from the initial guess $\tilde{\mu}$, take samples $\mathbf{y}_j$,

$j = 1, 2, \ldots, N$, from an importance density $f_1(\mathbf{y}, \tilde{\mu})$, and compute

$$\overline{M}_2(\mu_i) = \frac{1}{N}\sum_{j=1}^{N} \frac{g^2(\mathbf{y}_j) f^2(\mathbf{y}_j)}{f_1(\mathbf{y}_j, \mu_i) f_1(\mathbf{y}_j, \tilde{\mu})}, \quad i = 1, 2, \ldots, m \tag{20}$$

The new guess value for the next iteration $\tilde{\mu} = \mu_k$ at which $\overline{M}_2(\mu_k)$ in Eq. (20) is the minimum among m values of $\overline{M}_2(\mu_i)$, $i = 1, 2, \ldots, m$. Repeat this process until $\tilde{\mu}$ converges. Similar procedure can be applied to finding the optimal variance σ^2_{IS} of the importance density, or to obtaining both μ_{IS} and σ^2_{IS}.

3. Illustrative examples

In the section we attempt to illustrate the power of importance sampling techniques using two simple examples for flow and transport in one-dimensional porous media whose hydraulic properties are random constants rather than correlated random fields. One of the major reasons for choosing such simple examples is that analytical solutions for these examples are available, and therefore they can serve as the basis for comparing effectiveness of the conventional and IS-based Monte Carlo (ISMC) simulations. Another reason is that no clear strategy is yet available to optimally select importance density functions for correlated random fields, which shall be a topic of future research. Nevertheless, to our best knowledge these examples constitute first applications of importance sampling Monte Carlo techniques to subsurface flow and transport problems.

3.1. Head moments of 1D flow

In our first example, denoted as case 1, we consider a one-dimensional horizontal flow with prescribed deterministic influx q_0 on the left and prescribed deterministic constant head H_1 on the right. We assume that the log saturated hydraulic conductivity $Y = \ln K_s$ is a random constant, being a constant in the physical space but varying in the probability space with a normal distribution $N(\langle Y\rangle, \sigma^2_Y)$. For this case, as shown in Appendix A, the mean hydraulic head $\langle h(x)\rangle$ can be expressed as

$$\langle h(x)\rangle = H_1 + q_0(x - L)\mathrm{e}^{-\langle Y\rangle + \sigma^2_Y/2} \tag{21}$$

where x is the coordinate with $x = 0$ at the left boundary, and the head variance

$$\sigma^2_h = q_0^2(x - L)^2 \mathrm{e}^{-2\langle Y\rangle + \sigma^2_Y}(\mathrm{e}^{\sigma^2_Y} - 1). \tag{22}$$

Note that Eqs. (21) and (22) are exact solutions because there is no approximation involved in deriving these solutions, and therefore they are considered as references for comparison between the CMC and ISMC approaches. The parameters in our example are given as: $q_0 = 0.1$ m/day, $H_1 = 8.0$ m, $L = 100.0$ m, $\langle Y\rangle = -3.0$, i.e., $K_G = 0.04979$ m/day, and $\sigma^2_Y = 2.0$.

First, we need to find two pairs of optimal means and variances, one for importance sampling function $f_1(y)$ and the other for $f_2(y)$. We take samples Y_i, $i = 1, 2, \ldots, N$, from a distribution $N(\langle Y_{\mathrm{IS}}\rangle, \sigma^2_{Y,\mathrm{IS}})$, calculate $g(x, Y_i)$ according to $g(x, Y_i) = h(x, Y_i) = H_1 + q_0(x - L)\exp(-Y_i)$, and then compute sample variance $S^2_{H,N}$ and $S^2_{h^2,\mathrm{IS}}$ using Eqs. (10) and (18), respectively. This procedure is repeated for different values of $\langle Y_{\mathrm{IS}}\rangle = \langle Y_j\rangle$, $j = 1, 2, \ldots, m$, and $\sigma^2_{Y,\mathrm{IS}} = \sigma^2_k$, $k = 1, 2, \ldots, n$. Find optimal parameters that minimize Eqs. (10) and (18), i.e., those parameters corresponding to minimum $S^2_{H,N}$ and $S^2_{h^2,\mathrm{IS}}$ among $m \times n$ values. The dependence of $S^2_{H,N}$ and $S^2_{h^2,\mathrm{IS}}$ on $\langle Y_{\mathrm{IS}}\rangle$ and $\sigma^2_{Y,\mathrm{IS}}$ is illustrated in Fig. 2(a) and (b), showing two contour maps for $S^2_{H,N}$ and $S^2_{h^2,\mathrm{IS}}$, respectively. It is found that the optimized importance density $f_1(y)$ is $N(-3.8, 3.0)$ and $f_2(y)$ is $N(-5.9, 4.7)$, both of which greatly deviate from the original distribution $N(-3.0, 2)$. It is understandable that the large variance in importance density functions of Y represents the need to enlarge the tail from the original distribution, i.e., the

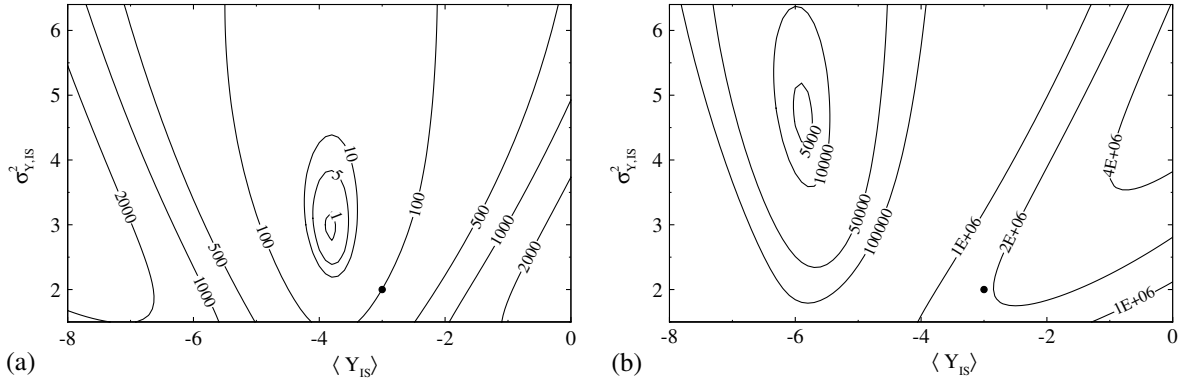

Fig. 2. Contour maps showing dependence of: (a) $S^2_{H,N}$ and (b) $S^2_{h^2,\mathrm{IS}}$ on parameters $\langle Y_{\mathrm{IS}}\rangle$ and $\sigma^2_{Y,\mathrm{IS}}$ of importance density functions. Optimal $\langle Y_{\mathrm{IS}}\rangle$ and $\sigma^2_{Y,\mathrm{IS}}$ are determined from the minimum $S^2_{H,N}$ and $S^2_{h^2,\mathrm{IS}}$ in the maps.

Y samples from the tails have larger impact on mean head than that of samples which are closer to the mean. While the left shift of the mean value, i.e., smaller $\langle Y_{IS} \rangle$, may imply that small values of Y have more "important" contribution to the mean head.

Figs. 3 and 4 compare mean head and head variance at the left boundary ($x = 0$) of the one-dimensional domain obtained from different approaches: the CMC simulation (CMC, 10^6 simulations), IS-based Monte Carlo simulation (ISMC, 10^5 simulations) for two sets of samples taken from $f_1(y)$ and $f_2(y)$, respectively. As a reference, the corresponding analytical solutions, i.e., Eqs. (21) and (22), are also shown in the figures. The figures show that ISMC simulation are several orders of magnitude more efficient than the CMC simulation. In addition, it seems that a smaller number of (IS-based) simulations is required for estimating mean head than for head variance. A few hundreds of simulations are enough for estimating the mean while thousands of simulations are needed for estimating the head variance. Furthermore, Fig. 4 indicates that for estimating the head variance the optimized importance density function $f_2(y)$ that minimizes $\sigma^2_{\bar{h}^2,IS}$ in equation (17) is more efficient than $f_1(y)$ that minimizes equation (8). Likewise, it is seen from Fig. 3 that $f_1(y)$ is more efficient than $f_2(y)$ in estimating the mean head. Figs. 5 and 6 convey the similar information as Figs. 3 and 4 but for the profiles of mean head and head variance in the 1D domain.

The performance of the CMC and ISMC approaches for estimating the mean head and the head variance at the left boundary for different levels of precision ϵ (here $\epsilon = 0.1$ means that the estimated values is within $\pm 10\%$ of the exact value) is tabulated in Table 1. The superior performance of the ISMC over the CMC is clearly evident. For example, in estimating the mean head at $\epsilon = 0.1$ and 0.05 levels, it takes just one run for the ISMC (using $f_1(y)$) while it takes 552 and 2342 runs for the CMC to achieve the same levels of precision. Similarly, in estimating the head variance, the ISMC (using

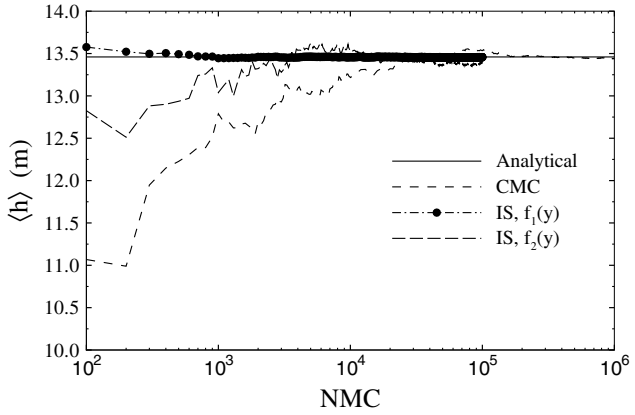

Fig. 3. Comparison of the rate of convergence for mean head at the left boundary obtained from different methods.

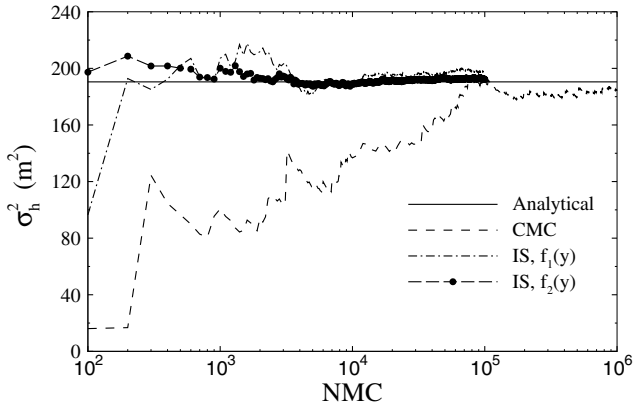

Fig. 4. Comparison of the rate of convergence for head variance at the left boundary obtained from different methods.

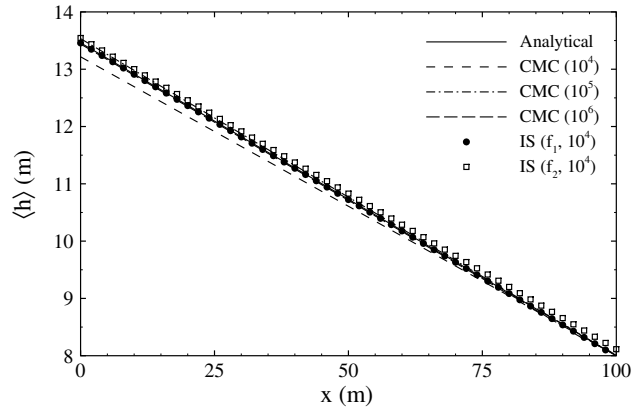

Fig. 5. Profiles of mean head obtained from analytical solution, importance sampling methods, and the CMC simulation with various sample sizes.

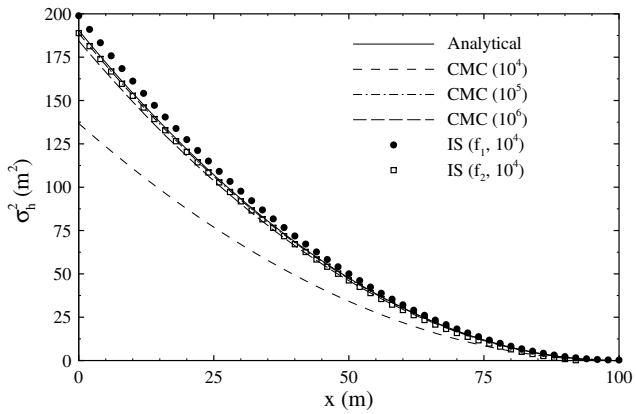

Fig. 6. Profiles of head variance obtained from analytical solution, importance sampling methods, and the CMC simulation with various sample sizes.

Table 1
Comparison of CMC and ISMC on the number of simulations required for different levels of precision

	$\langle h \rangle$			σ_h^2		
ϵ	0.1	0.05	0.01	0.1	0.05	0.01
CMC	552	2342	21,200	61,301	506,600	$>10^7$
IS, $f_1(x)$	1	1	79	2374	67,746	223,284
IS, $f_2(x)$	77	391	9647	387	1361	104,714

$f_2(y)$) is several orders of magnitude more efficient than the CMC for all levels of precision.

Although the Monte Carlo approach based on importance sampling converges much faster than the CMC approach, computational effort is needed in accurately determining the optimal importance densities. Here we investigated the effect of possible deviation from the optimal density. Figs. 7 and 8 show convergence of mean head and head variance at the left boundary resulted from ISMC simulations with importance densities $f_1(y)$ and $f_2(y)$ that are different from their optimal densities. Though simulations with the optimal density functions work much better (comparing to the analytical value) than those with density functions that are deviated from the optimal one, all of them are several orders of magnitude more efficient than the CMC simulation. This implies that it may not be necessarily to obtain a very accurate estimation of parameters of the importance density function and that the sampling intervals for constructing such contour maps as Fig. 2 can be relatively large, thus allowing great savings in computational efforts for determining suboptimal importance density functions.

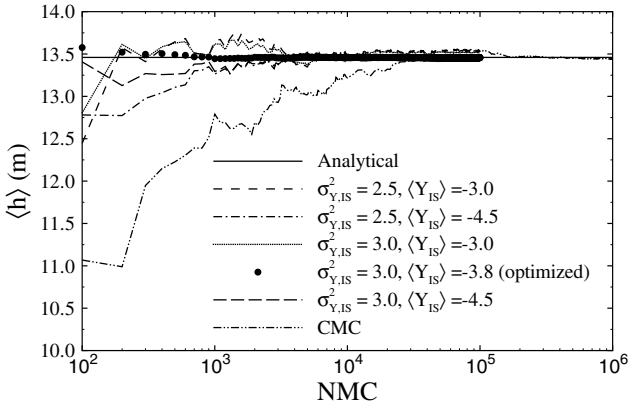

Fig. 7. Sensitivity of the importance sampling method on nonoptimal parameters for mean head simulation.

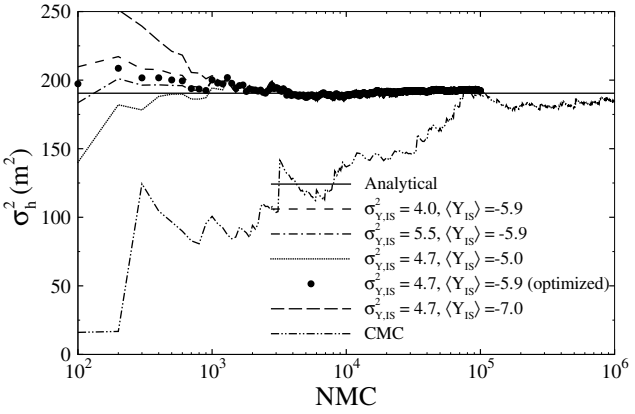

Fig. 8. Sensitivity of the importance sampling method on nonoptimal parameters for head variance simulation.

3.2. Particle travel time

In this example, we consider a case with a similar boundary configurations as in case 1, but here we assume that porosity ϕ is a random constant following a normal distribution $N(\mu_\phi, \sigma_\phi^2)$ and we are interested in the probability of particle's travel time less than a given value T. Since porosity ϕ is a random variable, so is travel time $t = L/(q/\phi) = L\phi/q$. In fact, the travel time $t \sim N(L\mu_\phi/q, L^2\sigma_\phi^2/q^2)$. The probability $P(t < T)$ can be written explicitly as

$$P_T = P(t < T) = P\left(\frac{L}{q}\phi < T\right) = P\left(\phi < \frac{Tq}{L}\right) = \Phi\left(\frac{Tq/L - \mu_\phi}{\sigma_\phi}\right) \tag{23}$$

where $\Phi(x)$ is the cumulative density function of the standard normal distribution. Eq. (23) is a basis in comparing effectiveness of the CMC and ISMC simulations.

Now if we want to estimate P_T using Monte Carlo simulation, we rewrite P_T as

$$P_T = P(t < T) = P\left(\phi < \frac{Tq}{L}\right) = \int_{-\infty}^{\frac{Tq}{L}} f(\phi)\,\mathrm{d}\phi = \int_{-\infty}^{\infty} g(\phi) f(\phi)\,\mathrm{d}\phi \tag{24}$$

where $f(\phi)$ is the pdf of variable ϕ, i.e., $N(\mu_\phi, \sigma_\phi^2)$, and $g(\phi)$ is a score function defined as $g(\phi) = 1$ if $\phi \leqslant Tq/L$, and 0 otherwise. Eq. (24) relates P_T to the mean of the score function $g(\phi)$ under density $f(\phi)$. The variance of $g(\phi)$ can be determined from

$$\sigma^2 = \int_{-\infty}^{\infty} [g(\phi) - P_T]^2 f(\phi)\,\mathrm{d}\phi = \int_{-\infty}^{\infty} g^2(\phi) f(\phi)\,\mathrm{d}\phi - P_T^2 = (1 - P_T)P_T \tag{25}$$

For the CMC simulation, one samples a sequence of ϕ_i, $i = 1, 2, \ldots, N$. If ϕ_i is less than or equal to Tq/L, then set $g(\phi_i)$ equal to unity. Otherwise, set $g(\phi_i)$ equal to zero, as illustrated in Fig. 9(a) for the case of $T = 250$ days. We collect the scores from all N samples and calculate the sample mean:

$$\overline{P}_{T,\mathrm{MC}} = \frac{1}{N}\sum_{i=1}^{N} g(\phi_i) \tag{26}$$

For given parameters $q = 0.1$ m/day, $\mu_\phi = 0.3$, $\sigma_\phi = 0.03$, i.e., $CV_\phi = 10\%$, and $L = 100.0$ m, Eq. (23) reduces to $P_T = \Phi(-10.0 + T/30)$. Table 2 gives P_T values obtained from the analytical solution, i.e., Eq. (23), for different T values. The table also gives the number of simulations required to estimate P_T using the CMC simulation (N_{MC}), within $\pm 10\%$ of the exact values.

From Table 2, it is seen that, with the decrease of parameter T (for T less than or equal to the mean travel time 300 days), the probability $P_T = P(t < T)$ decreases dramatically, and the number of realizations required to predict P_T using CMC increases significantly. For instance, to predict probability $P(t < 100$ days) with 10% accuracy will require over seven trillion realizations, which is clearly beyond capability of any computational scheme at the present time. The reason is that, for the given parameter values, the value of $f_t(t = 100$ days) or equivalently the value of $f(\phi = 0.10)$ is 1.31×10^{-11}, which is so small that it is almost impossible to take a sample ϕ from $f(\phi)$ such that $\phi < 0.1$. Another problem encountered in this kind of simulations is that currently it is difficult to generate such a large number of random numbers without repetitions. Although theoretically some algorithms, such as *ran*2() as presented in [12], have a period longer than 10^{18}, in a machine of 32-bit values of integers there are no more than $2^{31} = 2{,}147{,}483{,}647 < 10^{10}$ positive signed integers. Thus in a sequence of 10^{10} such integers there must be repetitions. It should be noted that the computational burden (the number of simulations required) will be even larger if we need more accurate predictions. For example, if we reduce $\epsilon = 0.1$ to 0.05, i.e., 5% precision, we have to increase the sample size by a factor of four.

Table 2
Number of simulations required for different values of T

T	P_T	N_{MC}	N_{MT}	N_{opt}
300	0.500	1.00×10^{2}	100	100
250	4.78×10^{-2}	1.99×10^{3}	202	151
200	4.23×10^{-4}	2.33×10^{5}	376	285
150	2.87×10^{-7}	3.49×10^{8}	567	437
100	1.31×10^{-11}	7.64×10^{12}	767	596
50	3.93×10^{-17}	2.54×10^{18}	970	759

Now instead of taking samples from $f(\phi)$, we use the importance sampling technique and take samples ϕ_i, $i = 1, 2, \ldots N$, from an importance pdf $f_1(\phi)$ of normal distribution $N(\mu_{\phi,\mathrm{IS}}, \sigma^2_{\phi,\mathrm{IS}})$. To preserve the mean, we have to define a new score function $g_1(\phi) = g(\phi)f(\phi)/f_1(\phi)$, which is shown in Fig. 9(b) for the cases of $T = 250$ days. Certainly, taking samples from $f_1(\phi)$ such that $\phi < 0.25$ is much easier than taking them from the original density $f(\phi)$, thus the sample size can be reduced.

After we sample ϕ_i, $i = 1, 2, \ldots, N$, from $f_1(\phi)$, we can estimate P_T by

$$\overline{P}_{T,\mathrm{IS}} = \frac{1}{N}\sum_{i=1}^{N} g_1(\phi_i) = \frac{1}{N}\sum_{i=1}^{N} g(\phi_i) f(\phi_i)/f_1(\phi_i) \tag{27}$$

The variance of the new score function $g_1(\phi)$ under distribution $f_1(\phi)$ can be derived as

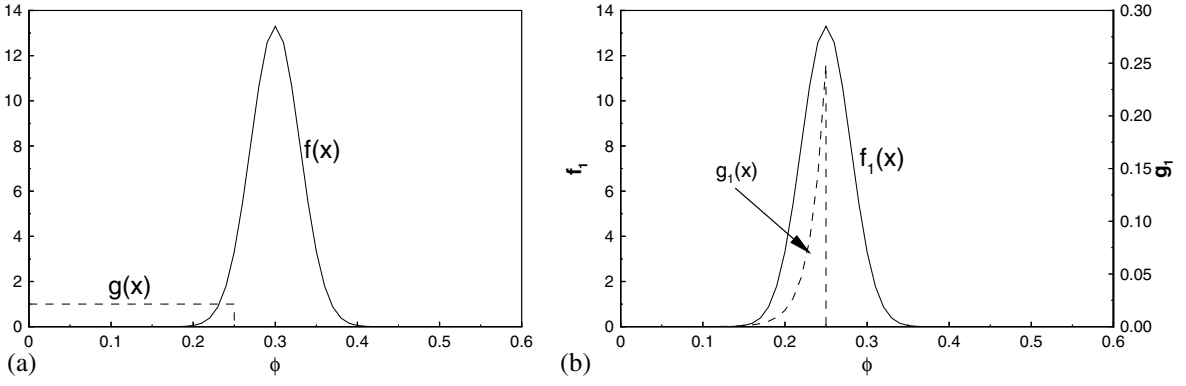

Fig. 9. Pdf and score function: (a) original, and (b) modified.

$$\sigma_{\mathrm{IS}}^2 = \int_{-\infty}^{\infty} [g_1(\phi) - P_T]^2 f_1(\phi)\,\mathrm{d}\phi = \frac{1}{\sqrt{A}} \frac{\sigma_{\phi,\mathrm{IS}}^2}{\sigma_\phi} \exp\left(\frac{B^2/A - C}{2\sigma_\phi^2 \sigma_{\phi,\mathrm{IS}}^2}\right) \Phi\left(\frac{\sqrt{A}(Tq/L + B/A)}{\sigma_\phi \sigma_{\phi,\mathrm{IS}}}\right) - P_T^2 \quad (28)$$

where $A = 2\sigma_{\phi,\mathrm{IS}}^2 - \sigma_\phi^2$, $B = \sigma_\phi^2 \mu_{\phi,\mathrm{IS}} - 2\sigma_{\phi,\mathrm{IS}}^2 \mu_\phi$, $C = 2\sigma_{\phi,\mathrm{IS}}^2 \mu_\phi^2 - \sigma_\phi^2 \mu_{\phi,\mathrm{IS}}^2$. The variance of the estimator $\overline{P}_{T,\mathrm{IS}}$ is σ_{IS}^2/N. In contrast to (25), the variance σ_{IS}^2 can be reduced by properly choosing $\mu_{\phi,\mathrm{IS}}$ and $\sigma_{\phi,\mathrm{IS}}^2$ of the importance density function $f_1(\phi)$. Notice that the expression for A implies $\sigma_{\phi,\mathrm{IS}}^2 \geqslant \sigma_\phi^2/2$.

Theoretically, the optimal density function can be determined by finding $\mu_{\phi,\mathrm{IS}}$ and $\sigma_{\phi,\mathrm{IS}}^2$ such that σ_{IS}^2 in Eq. (28) is minimized. However, it is not easy to directly minimize σ_{IS}^2 in Eq. (28). We utilize the large deviation theory (LDT) and find that that the optimized importance density is a Gaussian distribution $N(Tq/L, \sigma_\phi^2)$, i.e., an MT from the original density $N(\mu_\phi, \sigma_\phi^2)$, which is consistent with a previous result that for the Gaussian distribution, the LDT approach is the same as MT [14]. To confirm this numerically, we computed a series of σ_{IS}^2 based on Eq. (28) with different values of $\mu_{\phi,\mathrm{IS}}$ and $\sigma_{\phi,\mathrm{IS}}^2$ for two particular values of $T = 100$ and 250 days. Thus the probability of travel time $t < T$ for these two T values corresponds to the probability of $\phi < Tq/L = 0.1$ and 0.25, respectively. Fig. 10(a) and (b) show the performance of importance density functions, where the position of the dot in the figure stands for the mean and standard deviation of the original density $f(\phi)$. The measure of performance in Fig. 10 is represented by contour maps of $\log(\gamma) = \log(\sigma_{\mathrm{MC}}^2/\sigma_{\mathrm{IS}}^2) = \log(N_{\mathrm{MC}}/N_{\mathrm{IS}}) = \log(N_{\mathrm{MC}}) - \log(N_{\mathrm{IS}})$, showing the difference of the order of magnitude in the number of Monte Carlo simulations required for the CMC simulation and ISMC simulation. For instance, the curve for the contour level $\log(\gamma) = 2$ means that the number of simulations required for the CMC simulation is two orders larger than the number required for the ISMC simulation.

Several observations can be made from Fig. 10. First of all, for any given T and any fixed variance $\sigma_{\phi,\mathrm{IS}}^2$, sampling from distribution $N(Tq/L, \sigma_{\phi,\mathrm{IS}}^2)$ is most efficient, i.e., $\mu_{\phi,\mathrm{IS}} = Tq/L$ as predicted by the LDT. Notice that the optimal distribution is not $N(Tq/L, \sigma_\phi^2)$ but a slightly scaled Gaussian distribution with a smaller variance $\sigma_{\phi,\mathrm{IS}}^2 < \sigma_\phi^2$. Intuitively, this is understandable that while moving the mean of the sampling density to the dominating point which has the highest probability density (in the original density function) among other points in the domain of interest, D, the efficiency is further increased by "concentrating" the sampling distribution in the region around the dominating point. In addition, when sampling distribution is overbiased, the computational effort may increase rather than decrease, by many orders in worst cases. Thirdly, when the sampling mean is deviated form the optimized mean, the loss of efficiency may be compensated by increasing sampling variance $\sigma_{\phi,\mathrm{IS}}^2$. Furthermore, the efficiency that the IS techniques can achieve depends on the difference between μ_ϕ and the optimized mean $\mu_{\phi,\mathrm{opt}} = Tq/L$. For example, for $T = 100$ days, as illustrated in Fig. 10(a), the optimized importance sampling can achieve about 10 orders of magnitude more efficient than the CMC simulation. While for $T = 250$ days, the efficiency gain that the importance sampling can achieve is only about 1 order of magnitude. In the extreme case of $T = 300$ days, $\mu_{\phi,\mathrm{opt}} = Tq/L = 0.3 \equiv \mu_\phi$, the ISMC simulation does not have any efficiency gain over the CMC simulation. In this case, any density that deviates from $f(\phi)$ requires more computational effort.

Although the importance density derived from MT is slightly different from the optimal density, they are in

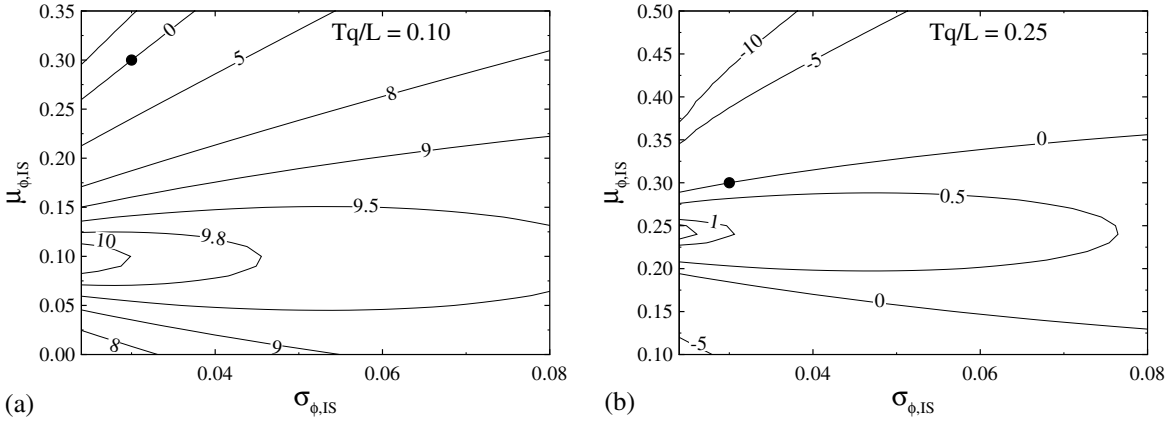

Fig. 10. Contour maps showing dependence of efficiency of importance techniques, presented by $\log(N_{\mathrm{MC}}/N_{\mathrm{IS}})$, on parameters $\langle\mu_{\phi,\mathrm{IS}}\rangle$ and $\sigma_{\phi,\mathrm{IS}}^2$ of importance density functions for (a) $T = 100$ days and (b) 250 days. Optimal $\langle\mu_{\phi,\mathrm{IS}}\rangle$ and $\sigma_{\phi,\mathrm{IS}}^2$ are determined from the maximum $\log(N_{\mathrm{MC}}/N_{\mathrm{IS}})$ in the maps.

Table 3
Comparison of P_T resulted from different approaches

T	P_T	$P_{T,\mathrm{MC}}$ (N_{MC})	$P_{T,\mathrm{IS}}$ (N_{IS})
300	0.500	0.500	0.500
250	4.78×10^{-2}	5.05×10^{-2} (2000)	4.62×10^{-2} (1000)
200	4.23×10^{-4}	4.53×10^{-4} (300,000)	4.33×10^{-4} (1000)
150	2.87×10^{-7}	–	2.64×10^{-7} (1000)
100	1.31×10^{-11}	–	1.20×10^{-11} (1000)
50	3.93×10^{-17}	–	3.58×10^{-17} (1000)

the same order of magnitude of efficiency compared to the CMC simulation, as also shown in Table 2 for two ISMC simulations with density functions derived from MT (N_{MT}) and from optimization (N_{opt}).

Numerical results from the CMC and ISMC simulations are tabulated in Table 3, where the number in parenthesis is the number of simulations used to compute P_T. As expected, for accuracy of 10%, 1000 simulations is enough for the ISMC approach for T value as low as 50 days, while a significant large number of runs is required for the CMC simulation to achieve the same level of accuracy even for $T = 200$ days. For $T < 150$ days, we are not able to obtain a good estimation of P_T using CMC.

4. Summary and conclusions

In this study we introduced a new approach, importance sampling techniques, and demonstrated its usefulness in simulating flow and transport in random porous media through two simple one-dimensional flow scenarios. The study leads to the following conclusions:

(1) In simulating flow and solute transport in random porous media, it is possible to save computational effort tremendously by employing importance sampling techniques. Because different random realizations may have a different impact on the quantities being estimated, realizations are taken from biased density functions (importance density functions) rather than the underlying original density functions such that those realizations with more "important" impact on estimated quantities are sampled more frequently. The effect of biased density functions is then corrected by weighting simulation outputs such that the means of the quantities being estimated are preserved. By choosing appropriate importance density functions, the ISMC simulation can be many orders of magnitude more efficient than the CMC simulation. In some cases, using ISMC approach, one can solve some problems that cannot be solved using the CMC simulation.

(2) The importance density for the mean estimator is chosen in such a way that the variance of the mean estimator is reduced while the mean is preserved. Therefore, the density function that minimizes the variance of the mean estimator does not necessarily minimize the variance of the variance estimator. In other words, in estimating both mean quantity and its associated prediction variance, one may need to find different importance density functions.

(3) In practice, we may not need to find the optimal importance function, an importance function that is relatively close to the optimal one in general works very well.

(4) Finally, we have emphasize that when the sampling distribution is overbiased, the computational effort may increase rather than decrease, by many orders in worst cases.

In examples presented in this paper, we assume that medium properties are random constant. To realistically apply importance sampling techniques to the real world, it is needed to extend this approach to correlated random fields.

Appendix A

For one-dimensional horizontal saturated flow with prescribed flux on the left and prescribed constant head on the right, in a porous medium whose log hydraulic conductivity is random constant, the governing equation for head h can be written as

$$\frac{\partial^2 h}{\partial x^2} = 0 \quad 0 \leqslant x \leqslant L \tag{A.1}$$

with boundary conditions $h = H_1$ at $x = L$ and $K_s \partial h/\partial x = q_0$ at $x = 0$. The exact solution for the above equation with boundary conditions is

$$h(x) = H_1 + q_0(x - L)\mathrm{e}^{-Y} \tag{A.2}$$

where $Y = \langle Y\rangle + Y'$. Taking ensemble mean of Eq. (A.2) yields the mean head

$$\langle h(x)\rangle = H_1 + q_0(x - L)\mathrm{e}^{-\langle Y\rangle + \sigma_Y^2/2} \tag{A.3}$$

Subtracting Eq. (A.3) from Eq. (A.2) gives

$$h'(x) = q_0(x - L)\mathrm{e}^{-\langle Y\rangle}\left[\mathrm{e}^{-Y'} - \mathrm{e}^{\sigma_Y^2/2}\right] \tag{A.4}$$

which leads to the expression for head variance

$$\sigma_h^2(x) = q_0^2(x - L)^2\mathrm{e}^{-2\langle Y\rangle}\mathrm{e}^{\sigma_Y^2}\left[\mathrm{e}^{\sigma_Y^2} - 1\right] \tag{A.5}$$

Here we used the results $\langle \exp(-Y')\rangle = \exp(\sigma_Y^2/2)$ and $\langle \exp(-2Y')\rangle = \exp(2\sigma_Y^2)$, which are exact for normally distributed Y'.

References

[1] Chen J, Lu D, Sadowsky JS, Yao K. On importance sampling in digital communications—Part I: Fundamentals. IEEE J Select Areas Commun 1993;11(3):289–99.

[2] Chen J, Sadowsky JS, Lu D, Yao K. On biasing Gaussian noise in importance sampling: optimization and implementation. In: Proceeding of SUPERCOMM/ICC '92: Discovering a New World of Communications, vol. 3, 1992. p. 1302–6.

[3] Dagan G. Stochastic modeling of groundwater flow by unconditional and conditional probabilities: 1. Conditional simulation and the direct problem. Water Resour Res 198;18(4):813–833.

[4] Dagan G. Flow and transport in porous formations. New York: Springer-Verlag; 1989.

[5] Davis BR. An improved importance sampling method for digital communication system simulation. IEEE Trans Commun 1986; COM-34:715–9.

[6] Gelhar LW. Stochastic subsurface hydrology. Englewood Cliffs, NJ: Prentice-Hall; 1993.

[7] Gelhar LW, Axness CL. Three-dimensional stochastic analysis of macrodispersion in aquifers. Water Resour Res 1983;19(1):161–80.

[8] Glynn PW, Iglehart DL. Importance sampling for stochastic simulations. Manage Sci 1989;35:1367–92.

[9] Lu D, Yao K. Improved importance sampling techniques for efficient simulation of digital communication systems. IEEE J Select Areas Commun 1988;6(1):67–75.

[10] Lu Z, Neuman SP, Guadagnini A, Tartakovsky DM. Conditional moment analysis of steady state unsaturated flow in bounded randomly heterogeneous porous soils. Water Resour Res 2002; 8(4):10.1029/2001WR000278.

[11] Neuman SP. Eulerian–Lagrangian theory of transport in space–time nonstationary velocity fields: Exact nonlocal formalism by conditional moments and weak approximations. Water Resour Res 1983;29(3):633–45.

[12] Press WH, Teukosky SA, Vetterling WT, Flannery BP. Numerical recipes in FORTRAN: the art of scientific computing. Cambridge University Press; 1992.

[13] Sadowsky JS. On the optimality and stability of exponential twisting in Monte Carlo estimation. IEEE Trans Inform Theory 1993;39:119–28.

[14] Smith PJ, Shafi M, Gao H. Quick simulation: a review of importance sampling techniques in communications systems. IEEE J Select Areas Commun 1997;15(4):597–613.

[15] Spanier J. An analytic approach to variance reduction. SIAM J Appl Math 1970;18:172–90.

[16] Tartakovsky DM, Neuman SP, Lu Z. Conditional stochastic averaging of steady state unsaturated flow by means of Kirchhoff transformation. Water Resour Res 1999;35(3):731–45.

[17] Yeh T-C, Gelhar LW, Gutjahr AL. Stochastic analysis of unsaturated flow in heterogeneous soils: 1. Statistically isotropic media. Water Resour Res 1985;21(4):447–56.

[18] Zhang D. Numerical solutions to statistical moment equations of groundwater flow in nonstationary, bounded heterogeneous media. Water Resour Res 1998;34:529–38.

[19] Zhang D. Stochastic methods for flow in porous media: coping with uncertainties. San Diego, California: Academic Press; 2002.

[20] Zhang D, Neuman SP. Eulerian–Lagrangian analysis of transport conditioned on hydraulic data: 1. Analytical–numerical approach. Water Resour Res 1995;31(1):39–51.

[21] Zhang D, Winter CL. Moment equation approach to single phase fluid flow in heterogeneous reservoirs. Soc Petrol Eng J 1999;4(2):118–27.